```
              GenCore version 5.1.3
       Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    February 11, 2003, 19:44:49 ; Search time 16.9714 Seconds
                                  (without alignments)
                                  2650.980 Million cell updates/sec

Title:         US-09-497-967-7
Perfect score: 2540
Sequence:      1 MKNNILVILIISLFINQIKS..........QCDFANFLSISLLLISYYLL 468

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters:    283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     PIR_73:*
               1: pir1:*
               2: pir2:*
               3: pir3:*
               4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	775.5	30.5	395	2	A46031	immobilization sur
2	256.5	10.1	677	2	C42125	trophozoite cystei
3	250	9.8	1766	2	A42125	trophozoite cystei
4	236.5	9.3	667	2	A48579	trophozoite surfac
5	228.5	9.0	713	2	A35502	major surface-labe
6	228	9.0	1274	2	T42017	cysteine rich prot
7	218.5	8.6	1372	2	T25933	hypothetical prote
8	216.5	8.5	596	2	A45664	variant-specific s
9	197	7.8	557	2	A48434	variant-specific s
10	191.5	7.5	1297	2	T30274	proteolialsin - se
11	190.5	7.5	2824	2	T22759	hypothetical prote
12	188.5	7.4	2823	2	T23064	hypothetical prote
13	188.5	7.4	2823	2	F87908	protein T22A3.8 [i
14	188.5	7.4	3102	2	T43291	laminin alpha chai
15	188	7.4	1609	1	MMHUB2	laminin gamma-1 ch
16	188	7.4	5376	2	T42215	zonadhesin - mouse
17	187.5	7.4	1459	2	T32271	hypothetical prote
18	183.5	7.2	1680	2	A43434	furin (EC 3.4.21.7
19	181.5	7.1	439	2	A36385	surface antigen se
20	179.5	7.1	1299	2	T43251	furin (EC 3.4.21.7
21	179.5	7.1	1548	2	S34583	serine proteinase
22	178.5	7.0	1607	1	MMMSB2	laminin gamma-1 ch
23	175.5	6.9	3075	2	S14458	laminin alpha-1 ch
24	174.5	6.9	3635	2	T10053	laminin alpha 5 ch
25	173	6.8	3712	2	S18253	laminin alpha-1 ch
26	172	6.8	2718	2	A23475	G surface protein
27	171	6.7	1895	2	T15881	hypothetical prote
28	169	6.7	1827	2	T34288	hypothetical prote
29	167.5	6.6	2704	2	S09118	G surface protein
30	167	6.6	2395	1	S50820	surface protein ty
31	166.5	6.6	969	1	A39490	subtilisin-like pr
32	166	6.5	3084	1	MMMSA	laminin alpha-1 ch
33	165	6.5	1700	2	S08167	Balbiani ring 3 pr
34	164	6.5	1797	2	T21888	hypothetical prote
35	164	6.5	1805	2	T21889	hypothetical prote
36	160.5	6.3	739	2	B88553	protein K04H4.2b [
37	160.5	6.3	2543	2	T31687	suface antigen - P
38	159	6.3	1252	2	S36016	oocyst wall protei
39	158	6.2	2533	2	T28675	alpha-51D immobili
40	158	6.2	2533	1	T28674	alpha-51D-immobili
41	158	6.2	3106	1	S53868	laminin alpha-2 ch
42	157	6.2	357	2	T21152	hypothetical prote
43	156	6.1	1391	2	T20406	hypothetical prote
44	156	6.1	1808	2	T15099	hypothetical prote
45	155.5	6.1	738	2	S40992	hypothetical prote

ALIGNMENTS

```
RESULT 1
A46031
immobilization surface I-antigen precursor - Ichthyophthirius multifiliis (fragment)
C;Species: Ichthyophthirius multifiliis
C;Date: 29-Sep-1999 #sequence_revision 29-Sep-1999 #text_change 07-Dec-1999
C;Accession: A46031
R;Clark, T.G.; McGraw, R.A.; Dickerson, H.W.
Proc. Natl. Acad. Sci. U.S.A. 89, 6363-6367, 1992
A;Title: Developmental expression of surface antigen genes in the parasitic ciliate I
A;Reference number: A46031; MUID:92335298; PMID:1631132
A;Accession: A46031
A;Molecule type: mRNA; protein
A;Residues: 1-395 <CLA>
A;Cross-references: GB:M92907; NID:g3628568; PIDN:AAC36158.1; PID:g3628569
A;Note: the authors translated the codon UUG for residue 330 as Ile
A;Note: sequence extracted from NCBI backbone (NCBIN:108734, NCBIP:108735); the seque
C;Genetics:
A;Genetic code: SGC5
C;Keywords: glycoprotein; surface antigen
F;2-395/Product: immobilization surface I-antigen #status experimental <MAT>
F;156,191,245,281/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match             30.5%;  Score 775.5;  DB 2;  Length 395;
Best Local Similarity   44.2%;  Pred. No. 3.7e-43;
Matches  175;  Conservative  35;  Mismatches 129;  Indels  57;  Gaps  16;

Qy    93 VKCPAGTAIAGGATDY-AAIITECVNCRINFY----------NENAP--NENAGASTCTA 139
            | :|||| || :|||| :||||:|:|||            |||||  ||| ||||||
Db     2 VPCPDGTQTOAGLTDVGAADLGTCVNCRPNFYYNGGAAQGEANGNQPFAANNAARGICVP 61

Qy   140 CPVNRVGGALTAGNAATIVAQCNVACPTGTALDDGVTTDYVRSFTECVKCRLNFYYNGNN 199
           :|||| |  || |: || |: ||||||||||||||| ||||||:| |||||||||
Db    62 CQINRVGSVTNAGDLATLATQCSTQCPTGTALDDGVTDVFDRSAAQCVKCKPNFYYNGGS 121

Qy   200 --GNTP-----FNPG----------KSQCTPCPAIKPANVAQATLGNDATITAQCNVACPDG 244
             ||:|     |||              |: ||  ||  | | |  | ||||:| ||| ||
Db   122 PQGEAPGVQVFAAGAAAAGVAAVTSQCVPCQINK--NDSPATAGAQANLATQCSNQCPTG 179

Qy   245 TISAAGVNNWVAQNTE--CTNCAPNFYNN-----NAPN---FNPG---------NSTC 283
             |:  || : | |||   | ||| ||||      |||    |||          :|||
Db   180 TVLDDGVT--LVFNTSATLCVKCRPNFYYNGGSPQGEAPGVQVFAAGAAAAGVAAVTSQC 237

Qy   284 LPCPANKDYGAEATAGGAATLAKQCNIACPDGTAIASGAT-NYVILQTECLNCAANFYFD 342
           :| || : ::  |||| :| |: |||:|||| | | | | : | :|||:| |||||||
Db   238 VPCQLNKN-DSPATAGAQANLATQCSTQCPTGTAIQDGVTLVFSNSSTQCSQCIANYFFN 296

Qy   343 GNNFQAGSSRCKACPANKVQGAVATAGGTATLIAQCALECPAGTVLTDGTTSTYKQAASE 402
           | : ||  ||  ||| :   |  :| ||:| ||    || ||| | |  ::||::|
Db   297 G-NLEAGKSQCLKCPVSKTTPAHA-PGNTATQATQCLTTCPAGTVLDDGTSTNFVASATE 354

Qy   403 CVKCAANFYTTKQTDWVAGIDTCTSCNKKLTSGAEA 438
           | :| |: |: : || : :| |:| |  |:|  |:
```

```
Db    355  CTKCSAGFFASKTTGFTAGTDTCTECTKKLTSGATA 390
```

RESULT 2

C42125

trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)

N;Alternate names: CRP72

C;Species: Giardia lamblia

C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 05-Dec-1998

C;Accession: C42125

R;Adam, R.D.; Yang, Y.M.; Nash, T.E.

Mol. Cell. Biol. 12, 1194-1201, 1992

A;Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170 ge

A;Reference number: A42125; MUID:92186850; PMID:1545800

A;Accession: C42125

A;Molecule type: DNA

A;Residues: 1-677 <ADA>

A;Cross-references: GB:M83934; NID:g159123

A;Experimental source: trophozoites

A;Note: sequence extracted from NCBI backbone (NCBIN:88443, NCBIP:88444); this ORF is no

C;Keywords: surface antigen

Query Match   10.1%;   Score 256.5;   DB 2;   Length 677;
Best Local Similarity   25.0%;   Pred. No. 2.3e-09;
Matches 121; Conservative 39; Mismatches 163; Indels 161; Gaps 32;

```
Qy   17  QIKSANCPVGTETNTAGQVDDLGTPANCVNCQKNFYINNAAAFV-----PGASTCTP---- 68
                : :|||||                            ::     :         ||
Db   50  QAAAAGC-----TKAGGAALDKMT-ATCEKCGDGYFLFMGGCYKTTDGPGSBICTKAEGG 103

Qy   69  -CPQKKDA-GAQPNPPATANLVTQCNVKCPAGTAIAGGATDYAAIITECVNCRINFYNEN 126
          |              | :     |  |||    :|:  :|:                |
Db  104  LCTECKTANGLFKNPAATPEKGSEC-ILC----SDINGADGYTG-VANCAQC-------T 150

Qy  127  APNFNAGASTCTACPVNRV---------------GGALTAGNAATIVAQCNVACPTGTALDDGVT 176
            ::|:||||||
Db  151  KSDSNKGAATCTACQAGYYKDFQACSKCDGTCLTCETSA---AQC-TSCPEG------- 198

Qy  177  TDYVRSFTECVKCRLNFYYNGNNGNTPFNPGKSQCTPCPAIKPANVAQATLGNDATITAQ 236
             |  ||      |:      |         ||||||            : |        |
Db  199  -KYLKGDKSCVN-----NNGCTGNTYADPESGKCLPCNTIDQA--------------CTQ 238

Qy  237  CNVACPDGTISAAGVNMWVAQNTECTNCAPNFYNNAPNFNPGNSTCLPCPANKDYGAEA 296
           |   ||   || :  |||||:                ||:  ||: :
Db  239  CEV---DST--------TKKPKCTNCGGQKMVKTAID--GTTTCV------------- 270

Qy  297  TAGGAATLAKQCNIACPDGTAIASGATNYVILQTECLNCAANFYFDGNNFQA--GSSRCK 354
                    |   ||    :    ||      ::  :::  :||       ||
Db  271  DANGCAT----SNV---DGSHFLNDGS------TKCILCSD----DSESLEANKGTPGCK 313

Qy  355  ACPANKVQGAVATA----------------GGTATLIAQCALECPAGTVLTDGTTSTYKQAASE 402
           |   |   |               ||       |::  :|: ::
Db  314  TCKKN---GAKPTCSECLDGYYNSGNGGTVTCEA-CGANC-----------ATCTQAGND 358

Qy  403  -CVKCAANFY----TTKQTDWVA------GIDTCTSCNK------KLTSGAEANLPESAKK 446
           | |:||||        ||: :|         |:|:||||       |:
Db  359  KCTKCKPGFFMKGNGPTGECVACDNAQGGIDGCAECTKSTGPLKCTKCKPNRKPAGTSD 418

Qy  447  NIQC 450
          ||
Db  419  NYTC 422
```

RESULT 3

A42125

trophozoite cysteine-rich surface antigen 170 - Giardia lamblia

N;Alternate names: CRP170; cysteine-rich surface antigen CRP170

C;Species: Giardia lamblia

C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 01-Dec-2000

C;Accession: A42125; B42125; S00530; S48056

R;Adam, R.D.; Yang, Y.M.; Nash, T.E.

Mol. Cell. Biol. 12, 1194-1201, 1992

A;Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170 ge

---

```
Db    355  CTKCSAGFFASKTTGFTAGTDTCTECTKKLTSGATA 390
```

A;Reference number: A42125; MUID:92186850; PMID:1545800

A;Accession: A42125

A;Molecule type: DNA

A;Residues: 1-98 <ADA1>

A;Cross-references: GB:M83937; NID:g159124

A;Experimental source: trophozoite

A;Note: sequence extracted from NCBI backbone (NCBIN:88421, NCBIP:88427); this ORF is

A;Note: the authors report but do not show 19 tandem repeats of the sequence of resid

A;Accession: B42125

A;Molecule type: DNA

A;Residues: 1269-1766 <ADA2>

A;Cross-references: GB:M83933; NID:g159122

A;Note: sequence extracted from NCBI backbone (NCBIN:88424, NCBIP:88431); this ORF is

J.Adam, R.D.; Aggarwal, A.; Lal, A.A.; de la Cruz, V.F.; McCutchan, T.; Nash, T.E.

J. Exp. Med. 167, 109-118, 1988

A;Title: Antigenic variation of a cysteine-rich protein in Giardia lamblia.

A;Reference number: S00530; MUID:88089405; PMID:3335828

A;Accession: S00530

A;Molecule type: DNA

A;Residues: 1154-1409,'A',1411-1420,'K',1422-1425,'R',1427-1481 <ADA3>

A;Cross-references: EMBL:X06741; NID:g9355; PID:g929603

R;Yang, Y.; Adam, R.D.

Nucleic Acids Res. 22, 2102-2108, 1994

A;Title: Allele-specific expression of a variant-specific surface protein (VSP) of Gi

A;Reference number: S48056; MUID:94301794; PMID:8029018

A;Accession: S48056

A;Molecule type: DNA

A;Residues: 1-56 <YAN>

A;Cross-references: EMBL:L25059

A;Experimental source: trophozoites WBA6

A;Note: the source is designated as Giardia intestinalis

C;Comment: This translation was produced by PIR staff from information provided by th

C;Genetics:

A;Gene: VSPA6

C;Keywords: surface antigen; tandem repeat

Query Match   9.8%;   Score 250;   DB 2;   Length 1766;
Best Local Similarity   24.3%;   Pred. No. 1.4e-08;
Matches 129; Conservative 44; Mismatches 175; Indels 182; Gaps 27;

```
Qy   24  PVGTETNTAGQVDDLGTPANCVNCQKNF-YYNNAAAFVPGASTCTCPQKKDAGAQPNPP 82
          |||||                      ::   |:|::::  ||||: :
Db   44  PYLKKTNPS---DPTGTCVSAVDCQGSAGYYTDDS--VSDAKECKKC--------NAP 88

Qy   83  ---ATANLVTQC----------------NVKCPAGTAIA-----GGA----TDYAAIITECV 116
           ||  |:||                          | |          |||:|||
Db   89  CTACAGTADKCTKCDANGAAPYLKKTNPSDPTGTCVSAVDCQGSAGYYTDDSVSDAKECK 148

Qy  117  NCRINFYNENAP-NFNAG-ASTCTACPVNRVGGALTAGNAATIVAQCNVACPTGT---AL 171
          ||           |||:  ||| |:|         :     ||:   |
Db  149  KC--------NAPCTACAGTADKCTKCDAN--------------GAAPYLKKTNPSDPTGTCVSAV 192

Qy  172  D-----DGVTTDYVRSFTECVKCRLNFYYNGNGNGNTPFNPGKSQCTPCPAIKPANVAQAT 226
           |       |||  :    ||||| |:  |           |: |::  :|:||
Db  193  DCQGSAGYYTDDSVSDAKECKKC------NAPCTACAGTADKCTKCDANGAAPYLKKT 244

Qy  227  LGNDATITAQCNVACPDGTISAAG--VNNWVAQNTECTNCAPNFYNNAPNFNPGNSTCL 284
               ||||          ::|:|  :   |:        |: ||
Db  245  NPSDPTGTCVSAVDCQ----GSAGYYTDDSVSDAKECKKC-----NAP---------CT 285

Qy  285  PCPANKDYGAEATAGGAATLAKQCNIACPDGTAIAS----GATNY---VILQTECLNC 335
          ||    |:  |   |||:                ::|:   ||: :   ::||  ||:|
Db  286  ACAGTADKCTKCDANGAAPYLKKTNPSDPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKC 345

Qy  336  AANFYFDGNNFQAGSSRCKACPANKVQGAVATAGGTATLIAQCALECPAGTVL------ 388
                  :  || :|: |:   |:      |: ||
Db  346  ---------------NAPCTACAGTADKCTKCDANGAAPYLKKTNPSDPTGTCVSAVDCQG 391

Qy  389  ---------TDGTTSTYKQ-------------AASECVKCAAN------ 409
                    |:||:: :|                ||||: ||||
Db  392  SAGYYTDDSVSDAKECKKCNAPCTACAGTADKCTKCDANGAAPYLKKTNPSDPTGTCVSA 451

Qy  410  ------FYTTKQTDWVAGIDTCTSCNKKLTSGAEANLPESAKKNIQCD 451
```

Db   452 VDCQGSAGYYT---DDSVSDAKECKKCNAPCTACA-----GTADKCTKCD 493

RESULT 4
A48579
trophozoite surface protein TSP11 - Giardia lamblia
C;Species: Giardia lamblia
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 29-Jan-1999
C;Accession: A48579
R;Ey, P.L.; Khanna, K.K.; Manning, P.A.; Mayrhofer, G.
Mol. Biochem. Parasitol. 58, 247-257, 1993
A;Title: A gene encoding a 69-kilodalton major surface protein of Giardia
A;Reference number: A48579; MUID:93241215; PMID:8479449
A;Contents: Ad-1
A;Accession: A48579
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-667 <EY1>
A;Cross-references: GB:M95814; NID:g159106; PID:g159107
A;Note: sequence extracted from NCBI backbone (NCBIN:130056, NCBIP:130058)

Query Match   9.3%;   Score 236.5;   DB 2;   Length 667;
Best Local Similarity 22.7%;   Pred. No. 4.5e-08;
Matches 128; Conservative 49; Mismatches 193; Indels 195; Gaps   30;

Qy   7 VILIISLFINQIKSANCPVGTETNTAGQVDDLGTPANCVNCQ-------------KNFY 52
            : ::| :: |:||| :: ||::|:|:|:::||:::   :||
Db   1 MLLAIFYFVISTLTAKCTTQTQTCEAEKCEMVGETEICTRCQTGKVPIDGKCVDATANAN 60

Qy   53 YNNAA-------AFVPGASTCTP-----CP-QKKDAGAQPNPPATANLVTQCNV 93
           ::| ||:|||| || || :| ||||| | ::||:|::|:|
Db   61 CKNASGDGDANQVCGKMSSVPGNTLCTTVSPDGVCSVAANEYFVPNADATHDSVVSCSE 120

Qy   94 KCPAGTA-------IAGGAT------DYAAIITECVNCRINFYNENAPNFNAGASTC 137
         ||||| | ||||| ||::||||||||||:|:||:||| ||:
Db 121 ETPIHLANNKQYIGVAGCATCSAPKAPGEDNTPKAATCTKCAAGFL--HTP--SEGLSSC 176

Qy 138 -TACPVNRVGGALTAGNAATIVAQCN-------CPDGTAIASGATNYVILQTECLNCAANEYFDGNNFQAGS 196
          ||||||||||||||||||||| ||:   :||:||||::|||||:||||||||||||| || : | ||
Db 138 -TACPVNRVGGALTAGNAATIVAQCNVACPTGTALDDGVTTDYVRSFTECVKCRLNFYY 196

Qy 177 EETCPEGYFGHTATAESKKT----CK-SCTGGSS----EAPNVRKGIGDCLKC-----MYN 222
            ||| :| || : || |:::| |::||| ||:| |
Db 177 EETCPEGYFGHTATAESKKT----CK-SCTGGSS----EAPNVRKGIGDCLKC-----MYN 222

Qy 197 GNNGNTPFNPGKSQCTPCPAIKPANVAQATLGNDAT--ITAQCNVA------CPDG---- 244
         ||| |:|::| |::||||||| |: :||: ||:|| ||:|||: |||:
Db 197 GNNGNTPFNPGKSQCTPCPAIKPANVAQATLGNDAT--ITAQCNVA------CPDG---- 244

Qy 223 EASGNT------LTCEKCSAQKKPSL-DKTSCNDCTGQNCAFCSSSGGDCEGCDSGFILD 275
                      :|:|||||| ::::| |||:|| || |:|:::: ||:|| ||::|||
Db 223 LTCEKCSAQKKPSL-DKTSCNDCTGQNCAFCSSSGGDCEGCDSGFILD 275

Qy 245 -------TISAAGVNNWVAQNTECTNCAPNFYNNAPNFNPGNSTCLPCPANKDYGA 294
                 |||||||||||||||| |||||||||||||||||||:||||||| | ||
Db 245 -------TISAAGVNNWVAQNTECTNCAPNFYNNAPNFNPGNSTCLPCPANKDYGA 294

Qy 276 GQNCVKSDCKTENCKACTNPKAANEVCTECI-----STHHLTPTSQCVYCQALGNYYA 329
         ||:| | ||::|| | |:|::|:| |||: ||:||:: ||| |: :||||
Db 276 GQNCVKSDCKTENCKACTNPKAANEVCTECI-----STHHLTPTSQCVYCQALGNYYA 329

Qy 295 EATAGGAATLAKQCNIA------CPDGTAIASGATNYVILQTECLNCAANEYFDGNNFQAGS 350
         |||||||||| | ||:| |||||||::||||||||||||||||||||| ||||| ||
Db 295 EATAGGAATLAKQCNIA------CPDGTAIASGATNYVILQTECLNCAANEYFDGNNFQAGS 350

Qy 330 GTNADNKKA-CKECTVANCKTCND------QGQCQTCNDGFYKNGDACSPCH 374
                  | ||:||:|||| | ||| ||:|::|||||||||||::| |
Db 330 GTNADNKKA-CKECTVANCKTCND------QGQCQTCNDGFYKNGDACSPCH 374

Qy 351 SRCKACPANKVQGAVATAGGTATLIAQCALECPAGTVL---TDGTTSTYKQ--------- 398
          ||| | ||::|||||||||||||||||||||||||| |||||||||
Db 351 SRCKACPANKVQGAVATAGGTATLIAQCALECPAGTVL---TDGTTSTYKQ--------- 398

Qy 375 ESCKTCSA---------GTA---SDCT-ECPTGKALKYGNDGTKGTCGEGCTTGQGSG 419
         |:||||:| ||| |||| ||||||||:|||||||||||:|||| ||:|
Db 375 ESCKTCSA---------GTA---SDCT-ECPTGKALKYGNDGTKGTCGEGCTTGQGSG 419

Qy 399 ---------AASECVKC---------AANFYTTKQTDWVAGIDTCTSC---- 428
                  ||||||| |||||||||||||||||||||||
Db 399 ---------AASECVKC---------AANFYTTKQTDWVAGIDTCTSC---- 428

Qy 420 ACKTCGLTIDGASYCSECDTQNEYPQNGICTSTTARTVATCKNSNVANGI--CSSCTNGF 477
         ||||| |:|||||||||| |:|| ||||||:|: | |:|:||||:||| | |:|||||
Db 420 ACKTCGLTIDGASYCSECDTQNEYPQNGICTSTTARTVATCKNSNVANGI--CSSCTNGF 477

Qy 429 ---------NKKLTSGAEAN 439
                  |||||:|:| |
Db 429 ---------NKKLTSGAEAN 439

Qy 478 LRMNGGCYETTKFPGKSVCEGANAD 502
         |||||||||||:|||||||||| |
Db 478 LRMNGGCYETTKFPGKSVCEGANAD 502

RESULT 5
A35502
major surface-labeled trophozoite antigen precursor - Giardia lamblia
C;Species: Giardia lamblia
C;Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 09-Sep-1997
C;Accession: A35502

R;Gillin, F.D.; Hagblom, P.; Harwood, J.; Aley, S.B.; Reiner, D.S.; McCaffery, M.; S.
Proc. Natl. Acad. Sci. U.S.A. 87, 4463-4467, 1990
A;Title: Isolation and expression of the gene for a major surface protein of Giardia
A;Reference number: A35502; MUID:90280395; PMID:2352929
A;Accession: A35502
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-713 <GIL>
A;Cross-references: GB:M33641; NID:g159131; PID:g159132
C;Keywords: surface antigen; transmembrane protein

Query Match   9.0%;   Score 228.5;   DB 2;   Length 713;
Best Local Similarity 24.4%;   Pred. No. 1.6e-07;
Matches 122; Conservative 34; Mismatches 170; Indels 175; Gaps   33;

Qy   23 CPVGTETNTAGQVDDLGTPANCVNCQKN-FYYN-----NAAAFVPGASTCTP----CPQ 71
         || || :|:|:|:|:||||||||||||: :||| |:||:|||| || | |
Db   69 CP---QHSAGK-------CTQCGGNSFMYKDGCYSSGEGLPGHSLCLSSDGDGVCTE 115

Qy   72 KKDAGAQPN---PPATANLVTQCNVKC--PAGTAIAGGATDYAAIITECVNCRINFYNEN 126
         |||||:||: ||||||||:|||: | |:| |:|||||||:||||||||||||||||||
Db 116 ----AAPGYFAPVGAAN-TEQSVIACGDTTGVTIAAGGNTYKG-IADCAEC------S 161

Qy 127 APNFNAGA-----STCTACPVNRVGGALTAGNAATIVAQCN------------------ 162
         ||||||| ||||:||||||||||||||||||:|||
Db 162 APDATAGAEAGKVATCTKCGVSK------YLKDNVCVDKAQCNSGSTNKFVAVDDSENGNKC 217

Qy 163 VACPT----GTALDDGVTTDYVRSFTECVKCRLNFYYNGNNGNTP------FNP- 206
         ||| | ||||||||:|||||||||||||||||||:|:||||| |||
Db 218 VSCSDNLNGGVANCDTCSYDEQSKKIKCTKCTDNNYLKTTSEGTSCVQKDQCKDGFFPKD 277

Qy 207 ---GKSQCTPCPAIKP--ANVAQATL-GNDATITAQCNVACPDG----------------- 244
            ||:||||||||| ||||||| ||:|||||||||||||:
Db 278 DSSAGNKCLPCNDSTDGIANCATCALVSGRSGAALVTCS-ACTDGYKPSADKTTCEAVSN 336

Qy 245 --TISAAGVNNWVAQNTECTNCAPNFYNNAPNFNPGNSTCLPCPA--NKDYGARATAGG 300
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db 337 CKTPGCKACSNEGKENEVCTDCDGSTY-----LTP-TSQCIDSCAKIGNYYGA-----TEG 386

Qy 301 AATLAKQCNIA---CPDGTAIASGATNYVLQTECLNCAANEYFDGNNFQAGSSRCKAC 356
         |||||||||:| |||||||:|||||||||||||||||||||||||||||||||:|||
Db 337 CKTPGCKACSNEGKENEVCTDCDGSTY-----LTP-TSQCIDSCAKIGNYYGA-----TEG 386

Qy 357 PANKVQGAVATAGGTATLIAQCALECPAGTVL----TDGTTSTYKQ--- 407
         |||||||||||||||||||||||||||||||| ||||||||||
Db 301 AATLAKQCNIA---CPDGTAIASGATNYVLQTECLNCAANEYFDGNNFQAGSSRCKAC 356

Qy 387 AKKLCKECTAANCKTCDD------------------QGQCQACNDGFYKNGDACSPCHESCKTC 432
         ||||||||||||||||| ||||||||||||||||||||||||||||
Db 387 AKKLCKECTAANCKTCDD------------------QGQCQACNDGFYKNGDACSPCHESCKTC 432

Qy 433 SA------GTA---SDCT-ECPTGKALRYGDDGTKGTCGEGCTTGTGAGACKTCG 477
         || ||| |||| ||||||||:||||||||||||:||||| |||:|||||
Db 357 PANKVQGAVATAGGTATLIAQCALECPAGTVL----TDGTTSTYKQ-----AASECVKCA 407

Qy 408 ANFYTTKQTDWVAGIDTCTSC 428
         |||||||||||||||||||||
Db 433 SA------GTA---SDCT-ECPTGKALRYGDDGTKGTCGEGCTTGTGAGACKTCG 477

Qy 478 LT------------IDGASYCSEC 489
         | ||||||||||
Db 408 ANFYTTKQTDWVAGIDTCTSC 428

Db 478 LT------------IDGASYCSEC 489

RESULT 6
T42017
cysteine rich protein - Giardia intestinalis
C;Species: Giardia intestinalis
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: T42017
R;Chen, N.; Upcroft, P.; Upcroft, J.
Parasitology 111, 423-431, 1995
A;Title: A Giardia duodenalis gene encoding a protein with multiple repeats of a toxi
A;Reference number: Z22027
A;Accession: T42017
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <CHE>
A;Cross-references: EMBL:L29079; NID:g951190; PID:g951191; PIDN:AAA74587.1
A;Experimental source: specific host: Homo sapiens
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

Query Match   9.0%;   Score 228;   DB 2;   Length 1274;
Best Local Similarity 23.1%;   Pred. No. 2.8e-07;
Matches 119; Conservative 35; Mismatches 183; Indels 178; Gaps   30;

```
Qy   21 ANCPV-GTETNTAGQVDDLGTPANCVNCQKNFYYN----------NAAAFVPGASTCTPC  69
Db  801 APCNVEGCETCVEGNAQ------QCKTCRPGYTINTDTKQCTKDPEAPCNVEGCETCV-  852

Qy   70 PQKKDAGAQP----NPPATANLVTQ------CNVK-CPAGTAIAGGATDYAAIITEC  115
Db  853 ----EGNAQQCKTCRPGYTINTDTKQCTKDPEAPCNVEGCE-TCVEGNA--------QQC  899

Qy  116 VNCRINFYNENAPNFNAGASTCTACPVNRVGGALTAGNAATIVAQCNVACPTGTALDDGV  175
Db  900 KTCRPGY-----TINTDTKQCTKDP---------------EAPCNV----------EGC  928

Qy  176 TTDYVRSFTECVKCRLNFYYNGNNG----------NTPFNPGKSQCTPCPAIKPANVAQA  225
Db  929 ETCVEGNAQQCKTCRPGYTINTDTKQCTKDPEAPCNTP------NCKTCDNPKTDNEI-C  981

Qy  226 TLGNDATITAQCNVACPDGTISAAGV----NNWVAQNTECTNCAPNFYNNNAPNFNPGN  280
Db  982 TKCNDGDYLTPTNQCVPDCT-AISGYYGDTDKKCKACNPECAECV----------GPAN 1029

Qy  281 STCLPCPANK-DYGAEATAGGAATLAKQCNIA-------CPD------GTAIASGATN--  324
Db 1030 NQCTACPVGKMLQYTDTNTPVNGGTCMDQCSVSSTNDGCAECGAQIGGTAYCSKCKNTQQ 1089

Qy  325 ---------YVILQTECLNCAANFYP-DG-------------NNFQAGSSRC  353
Db 1090 APLNGNCAASSRVAFCATTTSGACTKCNEGYFLKDGGCYQTDRQPGKQVCSNAQGGNGKC 1149

Qy  354 KACPANKVQGAVATAGGTATLIAQCALECPAGTVLTDGTTSTYKQAASECVKCAANFYTT  413
Db 1150 QTC--------ANGLAASDGNCA-ECHS----TCATCST-ADAADKCKTCATGYYKE 1192

Qy  414 KQTDWVAGIDTCTSCNKKLTSGAEANLPESAKKNI  448
Db 1193 NGDDTTAGL--CKKCSEKI-SGCKQCVSSSGSSVI 1224
```

RESULT 7

T25933
hypothetical protein W02C12.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T25933
R;Murray, J.; Wohldmann, P.
R;submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid W02C12.
A;Reference number: Z20112
A;Accession: T25933
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1372 <MUR>
A;Cross-references: EMBL:U80815; PIDN:AAB37995.1; GSPDB:GN00022; CESP:W02C12
A;Experimental source: strain Bristol N2; clone W02C12
C;Genetics:
A;Gene: CESP:W02C12.1
A;Map position: 4
A;Introns: 29/1; 66/1; 774/2; 823/2; 1046/1; 1108/2; 1298/1

Query Match   8.6%; Score 218.5; DB 2; Length 1372;
Best Local Similarity 22.2%; Pred. No. 1.3e-06;
Matches 90; Conservative 35; Mismatches 155; Indels 125; Gaps 20;

```
Qy   79 PN-PPATANLVTQCNVKCPAGTAIAGGATDYAAIITECVNCRINFYNENAPNFNAGASTC  137
Db  905 PNGRPDVGSLKIKEEYLCQAGQVV--------VRDLCVPCAPGTYH-----SAATGEC  949

Qy  138 TACPVNR--------VGGALTAGNAATIVAQCNVACPTGTALDDGVTTDYVRSFT  184
Db  950 ELCPIGEYQPLTARTECFKCAPGQITASEGAISEGECKDNCPPGHQY-DSLTSD------ 1002

Qy  185 ECVKCRLNFYYNGNNGNTPFNPGKSQCTPCPAIKPANVAQATLGNDATITAQCNVACPDG  244
Db 1003 -CVTCGYGYYQP----------SAGAFECIPC------GIGKTTLSEFATSEDECRDECPDG 1047

Qy  245 -TISAAGV----------NNWVAQNTECTNCAPNFYNN----------NAPNFNPG-  279
Db 1048 EQLSASGVCQPCQIGTYRSRGENKKCVACPPGTTEAIMSTRREQCNTPKCKPGQFLVKE 1107

Qy  280 NSTCLPCPANKDYGAEATAGGAATLAKQCNIACPDGTAIASGATNYVILQTECLN---CA  336
Db 1108 TKNCQFCPRGTFQNEEQES----------TCKLCAPDHTTAAPGAT---AESQCFSTNQCA 1155

Qy  337 ANFYF----------DGNNFQAGSSRCKACPANKVQGAVATAGGTATLIAQCALECPAG  385
Db 1156 TGEYNCSWHANCIDLPDENDVPSYECRCK--PGYRGNGTHCTDA----------CNDFC--- 1202

Qy  386 TVLTDGTTSTYKQAASECV----------KCAANFYTTKQTDWVAGI  422
Db 1203 --LNDGICKKKNNIGNVECICKDHFSGDRCELRFQASNNKLWIATV 1245
```

RESULT 8

A45664
variant-specific surface protein VSP1267 - Giardia lamblia
C;Species: Giardia lamblia
C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 29-Jan-1999
C;Accession: A45664
R;Mowatt, M.R.; Aggarwal, A.; Nash, T.E.
Mol. Biochem. Parasitol. 49, 215-227, 1991
A;Title: Carboxy-terminal sequence conservation among variant-specific surface protei
A;Reference number: A45664; MUID:92131058; PMID:1775165
A;Accession: A45664
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-596 <MOW>
A;Cross-references: GB:M63966; NID:g159140; PID:g159141
A;Note: sequence extracted from NCBI backbone (NCBIN:77609, NCBIP:77610)

Query Match   8.5%; Score 216.5; DB 2; Length 596;
Best Local Similarity 22.9%; Pred. No. 7.9e-07;
Matches 111; Conservative 40; Mismatches 180; Indels 153; Gaps 29;

```
Qy   40 TPANCVNCQKNFYYNNAAAFVPGASTCTPCPQKKDAGAQPNPPATANLVTQC----------  91
Db  117 TAGVCGACKDGYYKNSDA--VATADSCIAC----EDANCATCGGAGENKCTKCIDGYFVGA  171

Qy   92 ------NVKCPAGTAIAGGATDYAAIITECVNCRINFYNENAPNFNAGASTCTACPVNRV  145
Db  172 TGNEGGCIKCDATT------GPNSYKG-VAGCAKC----------EKPK-NAGPAKCIEC  213

Qy  146 GGALTAGNAATIVAQCNVACPTGTALDDG---VTTDIVRSFTE-CVKCRLNFYYNGNNG  200
Db  214 -----AADYLKTEADEQTSCVSEAVCRRGKTHFPTTDSAGGNKKVCVSC----------  257

Qy  201 NTPFNPGKSQCTPCPAIKPANVAQATLGNDATITA----------QCNVACPDGTISAA  249
Db  258 GTTNNGGIENCGECTSKESA----ARAGTEITCTKCSSNNLSPLGDACLTDCPAGTYAVS  313

Qy  250 GVNNWVA----------QNTECTNCAPNF--YNNNAPNFNPGNSTCL-  284
Db  314 GDSGSGVCKPCHNTCAGCQTDDRETSCTACYPGYSLLYESNG-------------ATGRCVKCTGAFI  368

Qy  285 ----------PCPANKDYGAEATAGGAATLAKQCNIACPDG----------TAIASGATNYVILQT  330
Db  369 TNCADGQCTAN----------VGG---AKYCT-QCKDGYAPIDGICTAVAAAGRDVSVCTA  415

Qy  331 ---ECLNCAANF-YPDGNNFQA----------GSSRKCACPANKVQGAVAT-AGGTA----TLIA  376
Db  416 TGGKCTACTGNYALLSGGCYNTQTLPGKSVCKAV-ANSNDGKCKTCANGQAPDPATNFCP  474

Qy  377 QCALECPAGTVLTDGTTST---YKQAASECVKCAANFYTTKQTDWVAGIDTCTSCNKK  431
Db  475 LCDSTCAECSTKNDADACTKCFPGYYKTGNKCIKCTESSNNGKKID--GIPDCLSCEAP  531

Qy  432 LTSG  435
```

Db   532 INTG 535

```
     :|
```

RESULT 9
A48434
variant-specific surface protein - Giardia lamblia (strain GS/M)
C;Species: Giardia lamblia
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 07-May-1999
C;Accession: A48434
R;Nash, T.E.; Mowatt, M.R.
Mol. Biochem. Parasitol. 51, 219-228, 1992
A;Title: Characterization of a Giardia lamblia variant-specific surface protein (VSP) ge
A;Reference number: A48434; MUID:92244292; PMID:1574080
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-557 <NAS>
A;Cross-references: GB:M80480; NID:g159142; PID:g159143

Query Match   7.8%;   Score 197; DB 2;   Length 557;
Best Local Similarity   20.7%;   Pred. No. 1.4e-05;
Matches 110;   Conservative 58;   Mismatches 211;   Indels 152;   Gaps 25;

```
Qy     7 VILIISLFINQIKSANCPVGTETNTAGQVDDLGTPANCVNCQKNFYNN------------ 55
           : | :  |    |   :| :  :||      | :  ||::| :
Db     1 MFLLINCLIASTLAGACSTTQANCVAEKCEMVGETEICTCQCKQNVYPINGVCEAAASSNT 60

Qy    56 ----AAAFVPGASTCTPC------------------------------PQKKDAGA----- 77
              ||::     || |                               | ::|| |
Db    61 KCQKASADEASDQTCGKCLSTTFMYKGGCYDKTGNLGRIICKTEGSDAGKCGACNDEKGF 120

Qy    78 QPNPPATANLVTQCNVKC--PAGTAIAGGATDYAAIITECVCRINFYNENAPNFNAGAS 135
          || | | || || || |  ||  ||  ||||   |  |||| |  :| |||| |||:
Db   121 FDNPDA-ANNVDSC-ISCGDATGVTIPGSSTKTYKGVAGCAKC--TKPSQISENTGTKEA 176

Qy   136 TCTACPVNRVGGALTAGNAATIVAQCNVA--CPTGTALDDGVTTDYVRSFTECVKCRLNF 193
          | |||| |   | | ||:|   |||| |  ||:|||| |||:|  |  |  |   | ||
Db   177 TCTECNANLYLKAVSSPTSAT---SCVSAEDCKTGYF----PTTDTTDSKKKCLTC---- 225

Qy   194 YYNGNGNTPFNPGKSQCTPCPAIKPANVAQATL---------GNDATITAQCNVACPDGT 245
          :   ||: ||     |  |  ||  ||| ||  :       ||| |||| || |||||
Db   226 ------STADKGGIDGCSACELLPSTTRASTVLISCSACSTNNLSPLKNECMQDCPAGT 278

Qy   246 ISAAGV-------NNWVAQNTE--CTNCAP-----NFYNNNAPNF-----------NPGN 280
          ||| ||       ||  ||| |  || |||     ||| ||  ||           | ||
Db   279 YADSNVCKPCHTSCASCKGDNTESSCTACYPGSVLSYGTDNTKGTCIAECTGKYLENCAD 338

Qy   281 STCLPCPANKDYGAEATAG------GAATLAKQCNIACPDGTAIASGATNYVILQTECLNC 335
          | | ||||  |||  | |       | | |||| | |||||  |||||  | ||| | ||
Db   339 GQCTATIAGSKYCSKCKSGFVPVNGLCVSAETARRAAPPGST---PDKTNGV-----CTAC 390

Qy   336 AANFYFD-GNNFQA-----GSSRCKACPANK----VQGAVATAGGTATLIAQCALECPAGT 386
          ||||| | || |      |:::| ||||||    | :| ||  :| ||||| | ||||
Db   391 TEKYFLESGGCYQAEKFPGNTLCTTADAGKCTTCANGQDKDSNGS----------CPA-- 438

Qy   387 VLTDGTTSTYKQAASECVKCAANFY--------TTKQTDWVAGIDTCTSC 428
          | |||||| || ||  | | |||||        : | ||| ||| |  |
Db   439 -CPTNCASCAKDNTKTCNKCFSGYYLDTAKACKKCSETSGNIQGVENCISC 488
```

RESULT 10
T30274
proteolaisin - sea urchin (Strongylocentrotus purpuratus) (fragment)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C;Accession: T30274
R;Laidlaw, M.; Wessel, G.M.
Development 120, 1325-1333, 1994
A;Title: Cortical granule biogenesis is active throughout oogenesis in sea urchins.
A;Reference number: Z20803; MUID:94298531; PMID:8026340
A;Accession: T30274
A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA
A;Residues: 1-1297 <LAI>
A;Cross-references: EMBL:U57753; NID:g1373379; PID:g1373380; PIDN:AAB02256.1

Query Match   7.5%;   Score 191.5; DB 2;   Length 1297;
Best Local Similarity   21.2%;   Pred. No. 6.6e-05;
Matches 125;   Conservative 63;   Mismatches 192;   Indels 211;   Gaps 37;

```
Qy    21 ANCPVGTETNTAGQVDDLGT------------------------------PANCVNCQK 49
           |                                              :||      |
Db   499 SGCPPTIVTCPAGRI-DCGTNVCVGARCDGVSDCSNGQDESGCPPTIVTCPAGRIDCGT 557

Qy    50 NFYYNNAAAFVPGASTCTPCPQKKDAGAQPNPPATANLVTQCNVKCPAGTAIAGGATDYA 109
            ||                |  :|||| | |     ||| ||  |||   |
Db   558 N---YCVVGARCDGVSDCS--NGQDESGCPP-----TIVT-----CPAGRIDCG--TNYC 600

Qy   110 AIITEC---VNCRINFYNENAPNFNAGASTCTACPVNRV--GGALTAGNAATIVAQCN- 162
                   :||          :   |:  || ||||| |  || |||| ||  || |
Db   601 VVGARCDGVSDC---------SNGQDEIGCPPTIVTCPAGRVDCGNNYCVVGSKCDGVSDCSN 654

Qy   163 ----------VACPTGTALDDGVTTDYVRSFTECV--KCRLNFYYNGNNGNTPFNPGK 208
                    | |  | |  ||       :  | ||  ||     |  ||  | | ||
Db   655 GQDESGCPPTIVTCPPG-RIDCG--TDY------CVVGARC----------DGVSDCSNGQ 696

Qy   209 SQ--CTP----CPA------------------------IKPANVAQATLGNDATITAQCNVACPDGTISAA 249
          |:   |||    ||                        |      |  ||  ||  |||:|
Db   697 DEIGCPPTIVTCPAGRVDCGNNYCVVGSKCDGVSDCSNGQDESGCPPTTSACPEGRVDCG 756

Qy   250 GVNNWVAQNTEC---TNCAPNFYNNAPNFNPGNSTCLPCPANK-DYGAEATAGGAATLA 305
                 |   |       |||   |: :  | ||       ||  |    | | ||
Db   757 --NNYCVVGSKCDGVSDCSNGQDESGCP----PTTST---CPEGRVDCGTDYCVVGA--- 804

Qy   306 KQCNIACPDGTAIASGATNYVILQTECLNCAANFYFDGNNFQAGSSRC--- 353
            || | || |  || |||  |   ||     | ||     ||||  |
Db   805 -RC-----DGVSDCSNGQDEIGCPPTIVTCPAGRVDCGNNYCVVGSKCDGVSDCSNGQDE 858

Qy   354 ----------KACPANKVQ--GAVATAGGTATLIAQCA----------LECPAGTV---- 387
                    |||||| |  || ||||| ||||| |          ||||| |
Db   859 SECPPTTSACPEGRVDCGNNYCVVGSKCDGVSDCSNGQDESGCPPTIVTCPAGRIDCGTN 918

Qy   388 ----------LTDGTTS-TYKQAASEC-----------------VKCAANFYTT-KQTDWVAGIDTC--- 425
                    | ||:||  ||||||:|                 ||||  || : |||| |   |
Db   919 YCVVGARCDGVSDCSNGQDESGCPPAIVTCPAGRVDCGNNYCVVGSKCD---GVSDCSNG 975

Qy   426 ----------TSCNKKLTS---GAEA--NLPESAKKNIQCDFANFLSISLLLI 463
                    |  :|    | ||||      | |     |    |    |  :  |||  |
Db   976 QDEEGCSFSSCRSRGDCEFFGTEAVESLEQRVTKLEQLDLENILVLENILV 1026
```

RESULT 11
T22759
hypothetical protein F55H12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C;Accession: T22759
R;Dobson, R.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19610
A;Accession: T22759
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2824 <WIL>
A;Cross-references: EMBL:Z81091; PIDN:CAB03143.1; GSPDB:GN00019; CESP:F55H12.3
A;Experimental source: clone F55H12
C;Genetics:
A;Gene: CESP:F55H12.3
A;Map position:
A;Introns: 87/1; 98/1; 126/2; 201/3; 343/3; 406/1; 576/3; 656/1; 825/3; 869/1; 909/1;
7/1; 1755/2; 1800/1; 1850/3; 1896/1; 2003/3; 2035/3; 2082/3; 2119/1; 2144/1; 2200/2;
C;Superfamily: LDL receptor ligand-binding repeat homology
F;243-279/Domain: LDL receptor ligand-binding repeat homology <LDL>

Query Match   7.5%;   Score 190.5;   DB 2;   Length 2824;

```
Best Local Similarity    21.9%;  Pred. No. 0.00016;
Matches  117; Conservative   42; Mismatches 207;  Indels 169; Gaps  25;

Qy     20  SANCPVGTET----NTAGQVDDLGTPANCVNCQKNFYYNNAAAFVPGASTCTPCPQKKD  74
               : :   : ||  :||||  |
Db   2123  SVKCMFGRQAKANNYNSTGSCKDRLPPTVLRCPKNIRKGTTEDFT----------  2168

Qy     75  AGAQPNPPATANLVTQCNVKCPAGTAIAGGATDYAAIITECVNCRINFYNENAPNFNAGA  134
                :
Db   2169  --------KVEWP---------DEDVAFFDNIGVIRIEVNYHNGQQFGVGI  2202

Qy    135  STCTACPVNRVGGALTAGNAATIVAQCNV--ACPT------GTALDDGVTTDYVRSFTE  185
               :|                           |  ::    |||:||:||:|||::|
Db   2203  TTV------RYVGFDAAGNSAECTFDVTIYQKSCPSQVYAEGGTVLAMQFTTAPFTAKAE  2256

Qy    186  CVKCRLNFY-----------YNGNGNTP--FNPGKSQCTPCPAIKPANVAQ  224
               |||  |                 :
Db   2257  KVKCDDNLYPTDSRPMFYVCDIMGDYQYGGWSDNVKQIYYLPACCQTS--PAVQAINGTV  2314

Qy    225  ATLGNDATI--------TAQCN--VACP---------DGTISAAG-VNNWVAQNTEC  261
               :|              :|  |  |  |        ||||:|:|
Db   2315  VGSSQCQQIHQRLRDVTWASADCDRILSCRLMIYPSCDEIDGRVSIADELSNIALQYTFS  2374

Qy    262  TNCA-------PNFYNNNAPNF---NPGNSTCLPCPANKDYGAEATAGGAATLAKQCNIAC  312
               |||               :|   |||||||||||||||||||||||||||||||||  |
Db   2375  TKNATETIDTTVLHNLQTNFTYIRQDSTVDCDP-----SYPIHDTNGNVTICVK------C  2424

Qy    313  PDGTAIASGATNYVILQTECLNCAANFYFDGNNFQAGSSRCKACPANKVQGAVATAGGTA  372
               ||:|| :|                ::|||||||||||||||||||||||||||||||
Db   2425  PEGT-FANKESN-----KCIDCPINTYRNSTNLD--QLKCTACPGTTVTGDVTGAVDE-  2474

Qy    373  TLIAQCALECPAGTVLTDG------------TTSTYK-------QAASECVKCAANFY  411
               :|||||||||||||||:||              |::|:        ||:|||||||||
Db   2475  ---SQCYVNCPIGQFESKGLCNCPCPEGTFGPTTGLRKCICCGFDLSTFGGPCIQCPRGLT  2531

Qy    412  TTKQT---DWVAGIDTCTTSCNKKLTSGAEEANLPESAKKNIQCDFANFLSIS  459
               |||:|    |     :                :                 |
Db   2532  TTSQASTSINSCDTTNCIDANTMINKNVTVGPSTPYSEIC---IACEQGTFQNVS  2583


RESULT 12
T23064
hypothetical protein T22A3.8 - Caenorhabditis elegans (fragment)
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C;Accession: T23064; T25096
R;Barlow, K.
A;Reference number: Z19669
submitted to the EMBL Data Library, October 1997
A;Accession: T23064
A;Status: preliminary
A;Molecule type: DNA
A;Cross-references: EMBL:AL008585; PIDN:CAA15432.1; GSPDB:GN00019; CESP:T22A3.8
A;Experimental source: clone H10E24
R;McMurray, A.
A;Reference number: Z19980
submitted to the EMBL Data Library, October 1996
A;Accession: T25096
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2823 <WI2>
A;Cross-references: EMBL:Z81125; PIDN:CAB03385.1; GSPDB:GN00019; CESP:T22A3.8
A;Experimental source: clone T22A3
C;Genetics:
A;Gene: CESP:T22A3.8
A;Map position: 1
A;Introns: 45/1; 282/2; 312/3; 416/2; 1255/3; 1329/3; 1418/3; 1776/2; 1988/2; 2760/2
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like H

Query Match     7.4%;  Score 188.5;  DB 2;  Length 2823;
Best Local Similarity   21.5%;  Pred. No. 0.00021;
Matches  103; Conservative   32; Mismatches 160;  Indels 185; Gaps  30;
```

```
Qy     22  NCPVGTETNTA----------------------------------GQVDDLG-------TP  41
               |||||:|||:|                                  ||||||
Db    860  NCPLGYEGNKCEYCSDGFFEDPLTGKCIECTCGNGNIDPMGIGNCDSETGKCLKCIGHTTG  919

Qy     42  ANCVNCQKNFYYNNAAAFVPGASTCTPCPQKKDAGAQPNPPATANLVTQC--NVKCPAG  98
               :|| |                ||:|||||||||  |     ||::|          :|
Db    920  DSCESC-KEHHWGNAQ---LHTCKPCGHTQGAVNP-----------QCSEENGECECK  963

Qy     99  TAIAGGATDYAAIITECVNCRINFYNENAPNFNAGASTCACPVNRVGGALTAGNAATIV  158
               :  |                    :||||:|||       ||||:|||:||
Db    964  ENYIG----AQCDRCK---ENHGDVENG---CPACDCNDTGS-IGSDCDQVS  1004

Qy    159  AQCNVACPTGTALDDGVTTDYVR---SFTE-CVKCRLNFYYNGNGNTPFNPGKSQC-  211
               :|||   |:  |         : |   :|||  |  |  |      :
Db   1005  GQCN--CKQGVF-GKQCDQCRPSYFNFTDAGCQFCHCNIYGSIEDGKCDQTTGKCECR  1059

Qy    212  -----TPCPAIKPA--NVAQA---TLGNDAT----------------ITAQCNVACPDGTISAAGV  251
                   :|       |  :|:|:    :|||                   ||:|  ||:|||:|:|:|
Db   1060  ENVEGTMCEKCADGYFNITSGDGCEDCGCDPTGSEDVSCNLVTGQC--VCKPG--------  1110

Qy    252  NNWVAQNTECTNCAPNFYNNNAPNFNPGNSTCLPCPANKDYGAEATAGGAATLAKQCNIA  311
               |      :||                 ||||||||||||||||||||||||||||||||
Db   1111  ----VTGLKCDSCLPNFYGLT----SEGCTECEPCP--------APGQVCPIDGSCV  1152

Qy    312  CPDGTAIASGATNYVILQTECLNCAANF--YFDGNNFQAGSSRCKACPANKVQGAVATAG  369
               ||:                   :|||||  ||||||||||||||||||||||||||||||
Db   1153  CPPNT---------VGEMCENCTNAWDYHPLNG--------CKILCDCSD---IGSDG  1190

Qy    370  GTA-TLIAQCALECPAGTVLTDGTTSTYKQAASECVKCAANFYTTKQTDWVAGIDTCTSC  428
               |:|                    |:|:|                |  |
Db   1191  GMCNTFTGQC--KCKAAYV---------------------GLKCDLCTHGFF------  1227
```

```
RESULT 13
F87908
protein T22A3.8 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-Aug-2001
C;Accession: F87908; E87908
R;anonymous; The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: F87908
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2823 <STO>
A;Cross-references: GB:chr_I; PIDN:CAA15432.1; PID:g3924779; GSPDB:GN00019; CESP:T22A
A;Accession: E87908
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2823 <ST2>
A;Cross-references: GB:chr_I; PIDN:CAB03385.1; PID:g3924881; GSPDB:GN00019; CESP:T22A
C;Genetics:
A;Gene: T22A3.8
A;Map position: 1
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-lik

Query Match     7.4%;       Score 188.5;  DB 2;  Length 2823;
Best Local Similarity   21.5%;  Pred. No. 0.00021;
Matches  103; Conservative   32; Mismatches 160;  Indels 185; Gaps  30;
```

```
Qy     22  NCPVGTETNTA----------------------------------GQVDDLG-------TP  41
               |||||:|||:|                                  ||||||
Db    860  NCPLGYEGNKCEYCSDGFFEDPLTGKCIECTCGNGNIDPMGIGNCDSETGKCLKCIGHTTG  919

Qy     42  ANCVNCQKNFYYNNAAAFVPGASTCTPCPQKKDAGAQPNPPATANLVTQC--NVKCPAG  98
               :|| |                ||:|||||||||  |     ||::|          :|
Db    920  DSCESC-KEHHWGNAQ---LHTCKPCGHTQGAVNP-----------QCSEENGECECK  963
```

```
Qy   99 TAIAGGATDYAAIITECVNCRINFYNENAPNFNAGASTCTACPVNRVGGALTAGNAATIV 158
          :|          ||
Db  964 ENYIG----------AQCDRCK----ENHGDVENG----CPACDCNDTGS--IGSDCDQVS 1004
Qy  159 AQCNVACPTGTALDDGVTTDYVR----SFTE--CVKCRLNFYYNGNNGNTPFNPGKSQC- 211
          :||              :|| ||
Db 1005 GQCN--CKQGVF--GKQCDQCRPSYFNFTDAGCQFCHCNIYGSIEDGKCDQTTGKCECR 1059
Qy  212 ----TPCPAIKPA--NVAQA---TLGNDAT---------ITAQCNVACPDGTISAAGV 251
          :|                |
Db 1060 ENVEGTMCEKCADGYFNITSGDGCEDCGCDPTGSEDVSCNLVTGQC--VCKPG------- 1110
Qy  252 NNWVAQNTECTNCAPNFYNNAPNFNPGNSTCLPCPANKDYGAEATAGGAATLAKQCNIA 311
          : | |||| ||||
Db 1111 ----VTGLKCDSCLPNFYGLT----SEGCTECEPCP------------APGQVCDPIDGSCV 1152
Qy  312 CPDGTAIASGATNVVILQTECLNCAANF--YFDGNNFQAGSSRCKACPANKVQGAVATAG 369
          :||                       ||
Db 1153 CPPNT-----------------------------VGEMCENCTTNAWDYHPLNG-------
Qy  370 GTA-TLIAQCALECPAGTVLTDGTTSTYKQAASECVKCAANFYTTKQTDWVAGIDTCTSC 428
                                         :||  ||
Db 1191 GMCNTFTGQC--KCKAAYV----------------------GLKCDLCTHGFF------- 1227
```

RESULT 15
MMHUB2
laminin gamma-1 chain precursor - human
N;Alternate names: laminin chain B2
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 10-Dec-1999
C;Accession: S13548; A28158; S13549; B34961; S14664; S23567
R;Kallunki, T.; Ikonen, J.; Chow, L.T.; Kallunki, P.; Tryggvason, K.
J. Biol. Chem. 266, 221-228, 1991
A;Title: Structure of the human laminin B2 chain gene reveals extensive divergence fr...
A;Reference number: S13548; MUID:91093128; PMID:1985895
A;Accession: S13548
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1609 <KAL>
A;Cross-references: GB:M55217; NID:g186937
A;Note: the nucleotide sequence was submitted to GenBank, February 1991
R;Pikkarainen, T.; Kallunki, T.; Tryggvason, K.
J. Biol. Chem. 263, 6751-6758, 1988
A;Title: Human laminin B2 chain. Comparison of the complete amino acid sequence with
A;Reference number: A28158; MUID:88198245; PMID:3360804
A;Accession: A28158
A;Molecule type: mRNA
A;Residues: 1-211,'R',213-1609 <PIK>
A;Cross-references: EMBL:J03202; NID:g186916; PIDN:AAA59488.1; PID:g307107
R;Fukushima, Y.; Pikkarainen, T.; Kallunki, T.; Eddy, R.L.; Byers, M.G.; Haley, L.L.;
Cytogenet. Cell Genet. 48, 137-141, 1988
A;Title: Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of the gen...
A;Reference number: S13549; MUID:89169663; PMID:3234037
A;Accession: S13549
A;Molecule type: mRNA
A;Residues: 1393-1609 <FUK>
A;Cross-references: EMBL:M27654; NID:g186923; PIDN:AAA59489.1; PID:g186924
R;Olsen, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasak...
Lab. Invest. 60, 772-782, 1989
A;Title: Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2 ...
A;Reference number: A34961; MUID:89280632; PMID:2733383
A;Accession: B34961
A;Molecule type: mRNA
A;Residues: 868-1551,'N',1553-1609 <OLS>
R;Santos, C.L.S.; Sabbaga, J.; Brentani, R.
DNA Seq. 1, 275-277, 1991
A;Title: Differences in human laminin B2 sequences.
A;Reference number: S14664; MUID:92216129; PMID:1806043
A;Accession: S14664
A;Molecule type: mRNA
A;Residues: 1282-1609 <SAN>
A;Cross-references: EMBL:X13939; NID:g34237; PIDN:CAA32122.1; PID:g34238
R;Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Pikkarainen, T.; Tryggvason, K...
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academi...
A;Reference number: S23567
A;Accession: S23567
A;Molecule type: DNA
A;Residues: 801-1481,'R',1483-1609 <VUO>
A;Note: mRNA was also sequenced
C;Genetics:
A;Gene: GDB:LAMC1; LAMB2
A;Cross-references: GDB:120136; OMIM:150290
A;Map position: 1q31-1q31
A;Introns: 140/1; 241/3; 285/2; 341/1; 404/1; 443/2; 476/2; 522/1; 563/1; 626/2; 664/
/3; 1525/1
C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin...
C;Function:

RESULT 14
T43291
laminin alpha chain - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
R;Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang...
submitted to the EMBL Data Library, June 1998
A;Reference number: Z22397
A;Accession: T43291
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3102 <ZHU>
A;Cross-references: EMBL:AF074902; PIDN:AAC26793.1
C;Genetics:
A;Map position: 1
A;Note: lama1/2
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h...

```
Query Match           7.4%;   Score 188.5;   DB 2;   Length 3102;
Best Local Similarity 21.5%;  Pred. No. 0.00023;
Matches 103;   Conservative 32;   Mismatches 160;   Indels 185;   Gaps 30;

Qy   22 NCPVGTETNTA------------------------GQVDDLG------------------TP 41
         ||   |                            |  ||
Db  860 NCPLGYEGNKCEYCSDGFFEDPLTGKCIECTCNGNIDPMGIGNCDSETGKCLKCIGHTG 919
Qy   42 ANCVNCQKNFYYNNAAAFVPGASTCTPCPQKKDAGAQPNPPATANLVTQC--NVKCPAG 98
          :||                                                ||
Db  920 DSCESC-KEHHWGNAQ-----LHTCKPCGHTQGAVNP--------QCSEENGECECK 963
Qy   99 TAIAGGATDYAAIITECVNCRINFYNENAPNFNAGASTCTACPVNRVGGALTAGNAATIV 158
          :|          ||
Db  964 ENYIG----------AQCDRCK----ENHGDVENG----CPACDCNDTGS--IGSDCDQVS 1004
Qy  159 AQCNVACPTGTALDDGVTTDYVR----SFTE--CVKCRLNFYYNGNNGNTPFNPGKSQC- 211
          :||              :|| ||
Db 1005 GQCN--CKQGVF--GKQCDQCRPSYFNFTDAGCQFCHCNIYGSIEDGKCDQTTGKCECR 1059
Qy  212 ----TPCPAIKPA--NVAQA---TLGNDAT---------ITAQCNVACPDGTISAAGV 251
          :|                |
Db 1060 ENVEGTMCEKCADGYFNITSGDGCEDCGCDPTGSEDVSCNLVTGQC--VCKPG------- 1110
Qy  252 NNWVAQNTECTNCAPNFYNNAPNFNPGNSTCLPCPANKDYGAEATAGGAATLAKQCNIA 311
          : | |||| ||||
Db 1111 ----VTGLKCDSCLPNFYGLT----SEGCTECEPCP------------APGQVCDPIDGSCV 1152
```

```
A;Description: interact with cells and with other basement membrane proteins to promote
C;Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C;Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-1609/Product: laminin gamma-1 chain #status predicted <MAT>
F;34-285/Domain: VI <DOM6>
F;286-504/Domain: V <DOM5>
F;286-339/Domain: laminin-type EGF-like homology <LE01>
F;342-395/Domain: laminin-type EGF-like homology <LE02>
F;398-442/Domain: laminin-type EGF-like homology <LE03>
F;445-492/Domain: laminin-type EGF-like homology <LE04>
F;495-504/Domain: laminin-type EGF-like homology #status atypical <LE05>
F;505-689/Domain: IV <DOM4>
F;690-1034/Domain: III <DOM3>
F;690-721/Domain: laminin-type EGF-like homology #status atypical <LE06>
F;724-770/Domain: laminin-type EGF-like homology <LE07>
F;773-825/Domain: laminin-type EGF-like homology <LE08>
F;828-881/Domain: laminin-type EGF-like homology <LE09>
F;884-932/Domain: laminin-type EGF-like homology <LE10>
F;935-980/Domain: laminin-type EGF-like homology <LE11>
F;983-1028/Domain: laminin-type EGF-like homology <LE12>
F;1035-1609/Domain: II/I <DOM1>
F;1035-1609/Region: heptad repeats
F;40-50/Disulfide bonds: #status predicted
F;60,134,576,650,1022,1107,1161,1175,1205,1223,1241,1380,1395,1439/Binding site: carbohy
F;1031,1034,1600/Disulfide bonds: interchain #status predicted

Query Match       7.4%;  Score 188;  DB 1;  Length 1609;
Best Local Similarity  23.9%;  Pred. No. 0.00014;
Matches  99;  Conservative  28;  Mismatches 165;  Indels 122;  Gaps  26;

Qy     75  AGAQPNPPATANLVTQCNVKCPAGTAIAGGATDYAAIITECVNCRINFYNENAPNFNAGA 134
               :  ! ! ! !     !  !   ! !     !!     ! ! ! !! !  !    ! !!
Db    674  ASARPGPGVPATWVESCT--CPVG---YGGQF------CEMC-LSGYRRETPNLGP-Y 718

Qy    135  STCTACPVNRVGGALTAGNAATIVAQC--NVACPTGTALDDGVTTDYVRSFTECVKCRLN 192
               ! !   ! :  !!   ! ! !   !  :           !! !    ! !!  !   !!
Db    719  SPCVLCACN---GHSETCDPETGVCNCRDNTAGP--------HCEKCSDG 757

Qy    193  FYYNGNNGNTPFNPGKSQCTPCP-------AIKP------ANVAQATLGNDATITAQCN 238
              :!   :  ! ! !  ! !     ! !       :  !!    !  !     !      :!
Db    758  YYGDSTAGTS------SDCQPCPCPGGSSCAVVPKTKEVCTNCPTGTTGK------RCE 805

Qy    239  VACPDGTISAAGVNNWVAQNTECTNCAPNFYNNNAPNFNPGNSTCLPCPAN-------- 289
            : ! ! !     !    !    !     ! !    ! !  !        !!
Db    806  L-CDDGYFGDPLGRNGPVRLCRLCQCSDNIDPNAVGNCNRLIGECLKCIYNTAGFYCDRC 864

Qy    290  KD-YGAEATAGGAATLAKQCNIAC-PDGTAIASGATNYYILQTECL---------NCAAN 338
               !  ! !     ! ! !  !    !! !    ! !  !  ! !!!
Db    865  KDGFFGNPLAPNPADKCKACN--CNPYGTMKQQSSCNPVTGQCBCLPHVTGQDCGACDPG 922

Qy    339  FYFDGNNFQAGSSRCKACPANKVQGAVATAGGTATL-IAQCALECPAGTVLITDGTTSTYK 397
            : ! ! !       !!  :    ! ! !  !      ! !    !! !
Db    923  FY---NLQSGQG-CBRCDCH----ALGSTNGQCDIRTGQC--ECQPGI-----------960

Qy    398  QAASECVKCAANFYTTKQTDWVAGIDTCTSCNKKLTSGAEANLPESAKKNIQCD 451
               :  ! !!    :    !    !  ! ! !    !   !  : !!  ! :! :
Db    961  -TGQHCERCEVNHFG------FGPEGCKPCD---CHPEGSLSLQCKDDGRCE 1002

Search completed: February 11, 2003, 19:48:47
Job time : 24.9714 secs
```